

November 21, 2024

Hon. Ona T. Wang
United States District Court
Southern District of New York

Re: *The New York Times Company v. Microsoft Corp.* No.: 23-cv-11195; *Daily News, LP v. Microsoft Corp.*: No. 1:24-cv-3285; *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.*: No. 1:24-cv-04872: Dispute Regarding Additional OpenAI ESI Custodians [ECF 314]

Dear Magistrate Judge Wang:

In what is now a monthly occurrence, Plaintiffs have moved to compel four new OpenAI custodians in addition to the eight they've previously requested. When combined with the 26 custodians OpenAI has already designated—who include multiple OpenAI executives and key members of its research, engineering, product and business teams—Plaintiffs' custodian wish list balloons to **38** different people. That figure equals roughly 25% of OpenAI's workforce in 2020 and 20% of OpenAI's workforce in 2021. But even that isn't enough. Plaintiffs have shown no indication that they intend to stop with their serial demands for additional custodians. Instead, they continue to demand invasive and duplicative discovery that would bury OpenAI in costly and burdensome document review. Plaintiffs' demands for additional custodians are unwarranted and disproportionate to the needs of the case. The Court should deny their motion.

I. Plaintiffs' motion is premature, inaccurate, and a waste of the Court's resources.

Plaintiffs first raised their latest demand for these four new OpenAI custodians on November 13. Pursuant to the ESI Order, OpenAI had until November 25 to respond. *See* Dkt. 246 at ¶ 5. Instead of waiting for OpenAI to investigate and respond, Plaintiffs filed this motion prematurely on November 19—four business days before OpenAI's deadline to accept, reject, or make a counterproposal with hit counts. *Id.* Nor did Plaintiffs wait for OpenAI to provide its hit counts and counterproposal for the 26 **existing** OpenAI custodians (due the very next day), which confirms that Plaintiffs' demands are overbroad and disproportionate to the needs of the case.

As such, in their haste to file, Plaintiffs provided the Court with inaccurate information about hit counts and the burden to OpenAI from further custodial review. For example, Plaintiffs claim that they have since “made a narrower proposal to yield around 500,000 documents for the current set of 26 agreed custodians” and “likely[]around 100,000 unique hits” for the eight custodians they moved to compel in August. ECF 314 at 3. But as OpenAI informed Plaintiffs on November 20, Plaintiffs' latest proposal actually **increased** the number of hits for OpenAI's existing 26 custodians from 632,357 to **659,407**. Further, the number of unique hits for the eight contested custodians that are the subject of Plaintiffs' August motion is 129,231—nearly 30% higher than what Plaintiffs' motion tells the Court. ECF 314 at 3. Neither of these figures

accounts for the four new custodians that are the subject of Plaintiffs' present motion. In short, Plaintiffs' demands for documents already far exceed the Court's 500,000 cap for OpenAI's current custodians; their demand for 12 total additional custodians only compounds the problem.

II. Plaintiffs should be ordered to accept OpenAI's proposed compromises as to Messrs. Lampe, Lightcap, Balaji, and Fedus.

Had Plaintiffs engaged in a meaningful meet and confer instead of rushing to file, they would have learned that OpenAI agrees to add Mr. Fedus, and that OpenAI was, and still is, open to compromises that avoid burdensome review for Messrs. Lampe and Lightcap. OpenAI's proposed compromise as to Mr. Balaji—which Plaintiffs rejected without explanation—will also ensure that OpenAI receives reciprocal discovery from Plaintiffs. Those compromises are as follows:

Michael Lampe. Plaintiffs rely heavily upon Mr. Lampe's designation as a custodian in the *Tremblay* case in the Northern District of California as a basis to add Mr. Lampe in the current action. But in doing so, Plaintiffs improperly suggest that the *Tremblay* Court made findings about Mr. Lampe's duties and relevance to the case. ECF 314 at 1. Not so. The language Plaintiffs quote is from the *Tremblay plaintiffs' argument*—not the Court. See *Tremblay v. OpenAI*, No. 23-cv-03223, ECF 191 at 2. Moreover, the *Tremblay* Court made clear that Mr. Lampe is one of the “final two of the 24 custodians” the *Tremblay* plaintiffs could designate, and “if Plaintiffs are choosing these custodians unwisely, they will have to bear the consequences of that decision.” *Id.* at 3. Plaintiffs here seek far more than 24 custodians. Nevertheless, OpenAI agrees to cross-produce in this action Mr. Lampe's custodial production from *Tremblay*.

Brad Lightcap. The *Tremblay* Court also recently ordered OpenAI to produce documents from Mr. Lightcap, identifying Mr. Lightcap (along with Mr. Lampe) as “the final two of the 24 custodians” designated in that case. *Id.* at 3. In the interest of compromise and efficiency, OpenAI agrees to cross-produce here Mr. Lightcap's custodial production from *Tremblay*.

Suchir Balaji. Plaintiffs seek former OpenAI employee Suchir Balaji as a custodian on the grounds that he (1) allegedly worked on post-training for ChatGPT and pretraining for the GPT-4 model; and (2) authored a blog post *after departing OpenAI* in which he shared personal thoughts on the copyright claims at issue in this case. Plaintiffs make no attempt to show that Mr. Balaji's documents are likely to corroborate the musings in his blog post. And the references to training are red herrings; OpenAI has already designated at least **nine** custodians on these topics.

Nevertheless, OpenAI offered to conduct a targeted search of Mr. Balaji's custodial file for internal complaints related to copyright concerns—provided that Plaintiffs agree to produce any documents they received from, and communications with, Mr. Balaji. Plaintiffs rejected that offer without explanation, because there is none. If the Court agrees Mr. Balaji's files are relevant, as Plaintiffs contend, then Plaintiffs should be ordered to produce them too. Plaintiffs' refusal to narrow the scope of Mr. Balaji's production and to provide reciprocal discovery is without merit. Their request to conduct a one-sided fishing expedition should be denied.

William Fedus. In the interest of compromise, OpenAI agrees to add Mr. Fedus as a custodian, subject to the Court's limitations on search terms and hit counts.

III. Plaintiffs' motions as to the remaining custodians should be denied.

Plaintiffs' demands for the remaining custodians are unjustified. First, as to Plaintiffs' new demands for Peter Deng and Ethan Stock, Plaintiffs fail to show they "would provide unique, relevant and noncumulative evidence." *See Coventry Cap. US LLC v. EEA Life Settlements, Inc.*, No. 17-cv-07417, 2021 WL 961750, at *2 (S.D.N.Y. Mar. 15, 2021); *see also* Hr'g. Tr. (Sep. 12, 2024) at 33:8-11; 101:20-23. Second, Plaintiffs' renewed demands for the same custodians it moved to compel on in August should again be denied. As to those custodians, Plaintiffs rely almost exclusively on documents produced **five months ago**; the issues have already been fully-briefed with the Court and illustrate the lack of evidence meriting reconsideration of Plaintiffs' arguments from August. OpenAI has produced over 31,000 documents in the last month. Plaintiffs should complete their review of those productions before making repetitive demands for more. *See* Hr'g. Tr. (Sep. 12, 2024) at 33:8-11; 101:20-23.

Peter Deng (new demand). Plaintiffs contend that their "continued review[of] OpenAI's document productions" has revealed justification for adding Mr. Deng as a custodian. ECF 314 at 1. But Plaintiffs' argument on Mr. Deng relies on a single produced document, produced to the Times in June – *i.e.*, months before Plaintiffs filed their initial motion to compel custodians. ECF 314-1 (Exhibit 4). Plaintiffs' repeated failure to review OpenAI's existing custodial productions before rushing to Court to demand more is improper. Moreover, the document confirms that Mr. Deng's information would be cumulative; it includes no less than four existing custodians.

Ethan Stock (new demand). Plaintiffs contend that Mr. Stock should be a custodian because he worked on [REDACTED]

ECF 314 at 2. But the exhibits Plaintiffs cite confirm he is duplicative. Exhibit 5, [REDACTED]. See ECF 314-1. Exhibit 6 [REDACTED]

[REDACTED]. See ECF 314-1.

Ilva Sutskever (subject of prior motion). As to Mr. Sutskever, Plaintiffs again rehash a document produced in June that was the subject of their prior motion. ECF 204-10. But Plaintiffs' prior motion did not explain why the document allegedly shows Mr. Sutskever has unique information, and Plaintiffs repeat that same failure here. The only new document Plaintiffs cite contains [REDACTED]

. See Exhibit 7. It further confirms he would be duplicative.

Wojciech Zaremba (subject of prior motion). Plaintiffs' argument as to Mr. Zaremba likewise cites documents produced more than five months ago, which were addressed in OpenAI's prior opposition. ECF 211 at 3. The only new exhibit Plaintiffs cite is a Slack that includes Mr. Zaremba and four existing custodians; it shows that at least two existing custodians are directly involved with and weighing in on the subject-matter at issue. Ex. 10.

* * *

For the foregoing reasons, Plaintiffs' motion should be denied.

Respectfully,

KEKER, VAN NEST &
PETERS LLP

/s Michelle Ybarra
Michelle Ybarra

LATHAM & WATKINS
LLP

/s Elana Nightingale Dawson
Elana Nightingale Dawson

MORRISON & FOERSTER
LLP

/s Carolyn Homer
Carolyn Homer

CC: All Counsel of Record (via ECF)

All parties whose electronic signatures are included herein have consented to the filing of this document.